BEBR

FACULTY WORKING PAPER NO. 1426

Asymptotic Theory and Econometric Practice

Roger Koenker

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 1426

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

January 1988

Asymptotic Theory and Econometric Practice

Roger Koenker, Professor
Department of Economics

ASYMPTOTIC THEORY AND ECONOMETRIC PRACTICE

Roger Koenker

University of Illinois at Urbana-Champaign

December, 1987

ABSTRACT

The classical paradigm of asymptotic theory employed in econometrics presumes that model dimensionality, p, is fixed as sample size, n, tends to infinity. Is this a plausible meta-model of econometric model building? To investigate this question empirically, several meta-models of cross-sectional wage equation models are estimated and it is concluded that in the wage-equation literature at least that p increases with n roughly like $n^{1/4}$, while that hypothesis of fixed model dimensionality of the classical asymptotic paradigm is decisively rejected. The recent theoretical literature on "large-p" asymptotics is then very briefly surveyed, and it is argued that a new paradigm for asymptotic theory has already emerged which explicitly permits p to grow with n. These results offer some guidance to econometric model builders in assessing the validity of standard asymptotic confidence regions and test statistics and may eventually yield useful correction factors to conventional test procedures when p is non-negligible relative to n.

Research supported by NSF Grants SES-8408567 and SES-8605595. A preliminary version of this paper was presented at the 5th World Congress of the Econometric Society in Cambridge, Mass., August 1985. The author wishes to express his thanks to S. Portnoy, A. Pagan, N. Keifer, L. MaGee, C. Manski, and G. Chamberlin for interesting conversations and/or correspondence on the subject of this paper. They are not accountable, of course, for any of the contents.

1. Introduction

The classical paradigm of asymptotic theory in econometrics rests on the following "willing suspension of disbelief." We must imagine a colleague in the throes of specifying an econometric model. Daily an extremely diligent research assistant arrives with buckets of (independent) new observations, but our imaginary colleague is so uninspired by curiosity and convinced of the validity of his original model, that each day he simply reestimates his initial model--without alteration--employing his ever-larger samples. Is this a plausible meta-model of econometric model building? Casual observation suggests that it is not. The parametric dimension of econometric models seems to expand inexorably as larger samples tempt the researcher to ask new questions and refine old ones. Indeed, this natural temptation is formally justified by the extensive literature on pre-testing and model selection. As larger samples improve the precision of our estimates, our willingness to accept bias in exchange for further improvements in precision inevitably declines. This viewpoint is quite explicit in the non-parametric regression literature for example.

In the next section we propose a simple, yet we hope plausible, meta-model of the econometric model specification process. And we present some empirical evidence on the specification of cross-sectional models of wage determination. We conclude from this exercise that the parametric dimension of wage models grows roughly like the fourth root of the sample size. The hypothesis of classical asymptotic theory that parametric dimension is fixed, i.e., independent of sample size, is decisively rejected. Should this crude empirical finding cause us to abandon our cherished beliefs in the consistency and asymptotic normality of econometric methods? Are the approximations suggested by fixed-p asymptotic theory "irrelevant" to the "real world" of econometric practice? In Section 3 we argue, on the contrary, that the forthright admission that $p \to \infty$ with n, offers an opportunity for a challenging and much more informative new form of asymptotic theory. We briefly review results of Huber (1973) on the large sample theory of the least squares estimator in linear models with

$p \to \infty$. Results of Yohai and Marrona (1979), Portnoy (1984,1985), and Welsh(1987) on large-p asymptotics for other M-estimators are then surveyed. It is hoped that this exercise will encourage others to think more critically about the dominant paradigm of asymptotic theory now employed in econometrics and contribute to the construction of a more realistic asymptotic paradigm.

2. Econometric Practice: A Meta-Model of Wage Determination Models

Models of wage determination offer an unusually rich and revealing source of data on the practice of model specification in econometrics. The "wage equation" pervades the applied econometrics literature: models of discrimination in employment, the effects of unions, returns to education, compensating differentials, etc. The development of several large scale panel surveys of labor market experience has facilitated the rapid growth of this empirical literature.

A meta-model is, of course, a model of models. As suggested in the previous section, we are primarily interested in modeling the dependence of the parametric dimension of models, say p, on the sample size of the available data, say n. Since the proposed dependent variable, p, is inherently a positive integer it is natural to begin with Poisson models in which the intensity (or rate) is taken to be some parametric function of the sample size and perhaps other characteristics of the research.

The data which we will analyze consists of 733 wage equations reported in 156 papers in mainstream economics journals and essay collections over the period 1970 to 1980. These papers deal with a variety of issues including returns to human capital, union effects discrimination, market structure effects, compensating differentials, etc. They are *all* cross-sectional models, and predominantly the cross-sectional unit is an individual, although in some cases it is some aggregate of individuals like a state, or industry. For each equation we observe the number of parameters estimated, the sample size, date of publication, and subject classified

into four categories. We also record the number of equations reported in each paper which is used to weight the observations. Inevitably, there are ambiguities in interpretation of the data. What constitutes an equation? Usually, this is quite straightforward, however, occasionally one finds samples split by age, race, sex, etc., and estimated with and without homogeneity constraints on the coefficients. Our policy in these cases was to interpret the disaggregated form of the equation as a single equation with say, mp, parameters, not as m distinct equations with p parameters. Frequently, there are non-wage equations in the surveyed papers; these are remorselessly ignored. Equations must have wage, or some function of wage as the dependent variable. Throughout, we have weighted observations on equations by the reciprocal of the number of equations appearing in the published paper. This tends to alleviate the problem of over-representation in the sample by a few (candid) "fishing" enthusiasts.

With the advent of the large panel datasets of labor economics, including census samples, some of the sampled wage-equations have exceedingly large sample sizes. A histogram of the meta-sample sample sizes is given in Figure 2.1. Since the horizontal scale is logarithmic in the figure, it is apparent that wage-equation sample sizes are roughly lognormally distributed.

It would be barbaric in the extreme to adopt a notation in which p was regressed on n, so we will revert to the more civilized convention of denoting our observed dependent variable by y, the sample size variable will be denoted z, and the vector of explanatory variables will be denoted x. Our meta-sample size, 733, may thus be denoted simply as n, and the dimension of x by p. This notational recursion makes the world safe for meta-meta-econometrics.

For the Poisson model we may write, for a typical observation

$$P(Y=y) = e^{-\lambda}\lambda^{y}/y!$$

while the rate parameter λ is expressed, e.g., as,

$$\lambda = \exp(x\beta) = \exp(\beta_1 + \beta_2 \log z)$$

In this form, the expectation and variance of the random variable Y are of course, both equal to the value λ. This is not entirely implausible since we might expect that the dispersion of model size would increase with its expectation. The Poisson hypothesis is obviously much stronger than this vague presumption of monotonicity and may be subjected to rigorous test. This problem is addressed explicitly below.

The first, simplest, and therefore perhaps the most compelling, of our estimated meta-models yields[1]

$$\log \lambda = \underset{(0.149)}{1.336} + \underset{(.017)}{0.235} \log z \tag{2.1}$$

Thus, roughly speaking, a 1% increase in the sample size of a wage determination model induces a 1/4% increase in the number of parameters of the model. This parsimony elasticity, or for the sake of brevity, "parsity," is, , *the* critical parameter of meta-econometrics. It will be denoted as π below. To put it slightly differently, p^4/n is roughly constant over the range of observed wage equation models. It must be emphasized that the maintained hypothesis of classical asymptotic theory that the dimension of parametric models is independent of sample size: $\beta_2 = 0$ in (2.1) is decisively rejected by the data. Unfortunately, our simple Poisson bivariate model is unsatisfactory in several respects:

1.) It predicts poorly for small n, implying negative degrees of freedom for n < 10 and extravagantly prodigal models for n < 100.

2.) The model, in GLIM parlance, is seriously overdispersed, i.e., the Poisson hypothesis that V(Y) = E(Y) is not supported by the data. The usual GLIM diagnostic is the estimated scale parameter

$$\tilde{\sigma}^2 = (n-p)^{-1}\Sigma(y_i-\hat{\lambda}_i)^2/\hat{\lambda}_i$$

is 4.73 in this case and significantly different from the hypothesized value of one.

3.) There are a few highly influential observations with z_i's (sample sizes) above 500,000.

[1] All estimation of Poisson models reported in this paper was carried out in the GLIM (Generalized Linear Interactive Modeling/System Release 3 Baker and Nelder (1978) see also McCullagh and Nelder (1983). Reported standard errors beneath the coefficients in all Poisson models are based on the GLIM quasi-likelihood model of McCullagh and Nelder(1983) in which $V(Y) = \sigma^2 E(Y)$ with σ^2 a free parameter, estimated as in point (2) below. If should be emphasized that in cases of overdispersion ($\sigma^2 > 1$) strict adherence to the Poisson assumption can seriously bias standard errors toward zero.

The narrow confidence interval on the coefficient of log z in (2.1) constructed conditional on this specification of the meta-model is far too optimistic. We have experimented with several alternate forms of the model. The obvious tactic of introducing a log quadratic term is (unfortunately) extremely sensitive to the observations alluded to in point (3.) above. With those observations, we obtain,

$$\log \lambda = \underset{(.512)}{-.438} + \underset{(.118)}{.663} \log z \underset{(.0067)}{-.0245} (\log z)^2 \tag{2.2}$$

while without them we have,

$$\log \lambda = \underset{(.512)}{1.737} + \underset{(.128)}{.0581} \log z + \underset{(.0078)}{.01543} (\log z)^2 \tag{2.3}$$

In the former the model predicts that model size declines after roughly n = 100,000, whereas the latter implies smoothly increasing parsity. In both cases parsity at mean[2] sample size ($n \approx 1000$) is roughly comparable to our simple model, $\pi = .32$ for (2.2) and $\pi = .27$ for (2.3). It is admittedly disturbing to find that the rise and fall of parsity is so sensitive to a few observations from our meta-sample. However, such sensitivity, especially in quadratic models, is often inevitable. Further, one may wish to question whether the observations with $n > 250{,}000$ are really drawn from the same population as the other observations of our meta-sample. For these cases, computational considerations enter the model specification process in a nontrivial way and may eventually come to dominate the "scientific" considerations which we emphasized in Section 1.[3] Thus we believe that there should be some *a priori* preference for (2.3) over (2.2).

Of the five subject categories which we have used to classify the papers only "discrimination" seems to have a significant (positive) effect. The others, "human capital", "unionism", and "women" are indistinguishable from the catch-all "general" category. Contrary to the

[2] Since sample sizes are logged this mean is geometric.

[3] This comment may seem to undercut our contention that $p \to \infty$ with n, which if taken absolutely literally is evidently asymptotically computationally infeasible. Of course, what is relevant is what happens in the range of practical experience which in the case of wage equations seems to be roughly sample sizes in the range 50-500,000. Here the evidence seems overwhelming that p increases gradually with n.

plausible hypothesis that increased computing power has led to bigger models over time, the inclusion of an explicit annual trend yields a *negative*, but insignificant, coefficient. Neither of these auxiliary subject or vintage variables have a substantive effect on the relationship between model dimension and sample size and they have been omitted from the reported models.

We have also experimented with models in log (log n). The estimated Poisson model

$$\log \lambda = \underset{(.315)}{-.777} + \underset{(.148)}{1.947} \log \log z \tag{2.4}$$

yields a slightly better fit than our simple meta-model (2.1) and at mean sample size it implies a parsity of $\pi = .28$. This "law of the iterated logarithm" form of the meta-model has the attractive feature that the parsity parameter is proportional to the reciprocal of log (sample size), and therefore tends to zero as $n \to \infty$ albeit slowly. Figure 2.2 illustrates the differences among the four models reported above with respect to parsity as a function of sample size. One sees clearly in the Figure that the differences between the functional forms are primarily in the extremes of the observed sample sizes.

We have emphasized above that all of the Poisson models suffer from over-dispersion, that is, the estimated conditional variance of dependent variable is considerably larger than the conditional mean that is predicted by the Poisson model. One interpretation of this over-dispersion in Poisson models is that there is some inherent variability in the rate parameter λ around its hypothesized (log) linear form. The classical approach to treating this (common) syndrome is to hypothesize a random intercept for the rate equation, with a gamma distribution and on integrating out this random coefficient one obtains a negative binomial model for the dependent variable. See Appendix A for details. This approach may be traced to Anscombe (1949) who applied it in entomology. A recent application in econometrics is Hausman, Hall, and Griliches (1983), and an extremely insightful view of this problem and parametric heterogeneity in general is provided by Chesher(1984), and Cox (1983).

Tests for parametric heterogeneity in Poisson models may be developed along the lines suggested by Lancaster (1984) based on Chesher (1984), White (1982), Cox (1984) and others. The basic information identity

$$D = E\nabla^2 \log f + E(\nabla \log f \nabla' \log f) = 0$$

and its extensions may be used to construct tests which are readily computed as nR^2 from a regression of a column of ones on a matrix of n by $p(p+1)/2$ elements of D augmented by the matrix of gradient "observations" $g = \nabla \log f$ evaluated at the maximum likelihood estimator. "Explanatory power" in this regression suggests systematic departures in the fitted model from the hypothesis that D and g have zero expectation. Several of these tests have been conducted restricting attention to the components of $[D \vdots g]$ corresponding to the intercept parameter in the log λ equation. Here the test is particularly simple since $\hat{d}_i = (y_i - \hat{\lambda}_i)^2 - \lambda_i$ and $\hat{g}_i = y_i - \hat{\lambda}_i$ where $\hat{\lambda}_i = \exp(x_i \hat{\beta})$. The test statistic is 133.1 for meta-model (2.1) for example, which is clearly an implausible value for a central χ^2 random variable on 2 degrees of freedom. In this context, this "White test" is closely related to the GLIM diagnostic referred to above, see Cameron and Trivedi(1985) for detailed discussion.

Unfortunately, the negative binomial model while quite attractive from a number of perspectives is somewhat unwieldy computationally. Estimation in GLIM may be carried out by conditioning on the variance parameter, but this approach yields unsatisfactory (conditional) estimates of standard errors. Some exploratory forays have been made using the negative binomial model and the remarkable quasi-maximum likelihood estimation software of Spady (1984). This approach is somewhat capital intensive, but avoids the labor of coding analytical derivatives, and has the virtue of producing statistically reliable standard errors.[4] In the simple loglog model we obtain

$$\log \alpha_i = \underset{(.365)}{-.679} + \underset{(.200)}{1.900}\text{loglog } z$$

[4] Standard errors are computed by numerical approximations to the general quasi-mle formula $V = J^{-1} I J^{-1}$ where I denotes $E \partial \log f / \partial\theta \partial \log f / \partial\theta'$ and J denotes $E \partial^2 \log f / \partial\theta\partial\theta'$.

with $\hat{\gamma}$ = 1.51 (.14). Here $EY_i = \alpha_i \gamma$ so the parsity parameter has the same interpretation as in the loglog Poisson model and it is somewhat comforting to observe that the results are essentially indistinguishable from that model.

3. Asymptotic Theory: A Practical Paradigm

We are thus faced with the familiar dialetical discrepancy between theory and practice. Theory offers us a static view of the econometric model, a model "cast in concrete," unperturbed by the influx of new data. The practice of econometrics, however, offers quite a different, more plastic, view: models gradually expanding and elaborating themselves in response to the availability of new data. How are these views to be reconciled?

The answer, of course, is to expand the paradigm of classical asymptotic theory. Huber (1973) was apparently the first to observe that, under rather mild regularity conditions on the sequence of designs, consistency and asymptotic normality of the least-squares estimator in linear models was possible if $p/n \to 0$. These results are quite elementary, on the same level as the fixed p asymptotics which are done in introductory graduate courses, and therefore should be better known. To my knowledge, only the recent text of Amemiya (1985) treats any of these questions.

To illustrate the general approach consider the simplest application to the classical linear model with iid disturbances: the asymptotic behavior of the least-squares estimator. For fixed p, and error distributions with finite variance, we know that $\hat{\beta} \to \beta_0$, strongly if and only if $(XX)^{-1} \to 0$. See Lai, Robbins and Wei (1979), for a proof of this surprisingly delicate result. For $p \to \infty$ with n, consider the "hat" matrix $H = X(X'X)^{-1}X'$ We know the following: $h_{ii} \in [0,1]$, $tr(H) = p$, $HH = H$ Thus, since $\hat{y} = Hy$, we have

$$Var(\hat{y}_i) = \sum_{k=l}^{n} h_{ik}^2 \sigma^2 = h_{ii} \sigma^2$$

so by Chebyshev's inequality

$$P[|\hat{y}_i - Ey_i| \geq \epsilon] \leq h_{ii}\frac{\sigma^2}{\epsilon^2} \tag{3.3}$$

Thus $\hat{y}_i \to^P x_i\beta$ if $h_{ii} \to 0$; the converse is also true, see Huber(1973). Note that $\bar{h} = \max_i h_{ii} \geq n^{-1}\sum h_{ii} = n^{-1}Tr(H) = p/n$, so $\bar{h} \to 0$ implies $p/n \to 0$ so $p/n \to 0$ is necessary, but not sufficient, for weak consistency.

Now consider an arbitrary linear function of $\hat{\beta}$, say $\alpha'\hat{\beta}$, $\|\alpha\| = 1$. Assume F isn't Gaussian, and reparameterize so that $X'X = I_p$ Hence, $\hat{\beta} = X'y$ and $a = \alpha'\hat{\beta} = \alpha'X'y \equiv s'y$ where $s's = \alpha'X'X\alpha = 1$ so $Var(a) = \sigma^2$ Then a straightforward applications of the Lindeberg Central Limit Theorem implies that a is asymptotically Gaussian if and only if $\max_i |s_i| \to 0$. Bickel(1977) has reformulated this as: estimable functions $\alpha'\hat{\beta}$, are asymptotically Gaussian with natural parameters if and only if the fitted values are consistent.

These results for the least squares estimator are extremely encouraging. What happens in nonlinear cases? The simplest nonlinear case is robust regression for linear models. Here all the nonlinearity seems to be very well circumscribed, however, already, serious difficulties arise. Huber (1973), on the basis of informal expansions and Monte Carlo experimentation conjectured that $p^2/n \to 0$ was necessary to achieve a uniform normal approximation for a typical M-estimator in the absence of any symmetry conditions on the error distribution. Subsequently, Yohai and Marrona (1979) showed that $p^{3/2}\bar{h} \to 0$ implied a uniform normal approximation, but this means, since $\bar{h} \sim p/n$, that $p^{5/2}$/n would be sufficient. Huber (1981) conjectured that $p\bar{h} \to 0$ was sufficient and that $\sqrt{p}\bar{h} \to 0$ was necessary if the error distribution was permitted to be asymmetric. For symmetric errors one might hope that $\bar{h} \to 0$ was sufficient as in the least-squares case. Huber (1980) contains an elementary proof for the case $p^2\bar{h} \to 0$.

Portnoy (1984, 1985) has substantially improved these results and verified an important conjecture of Huber. In particular, he shows that under quite mild regularity conditions on X, $p(\log n)/n \to 0$, suffices for norm consistency of M-estimators based on (smoothly) monotone ψ functions. Asymptotic normality is more problematic, and under slightly stronger regu-

larity conditions, Portnoy shows that if $(p\log p)^{3/2}/n \to 0$ then a uniform normal approximation is possible. Note that this essentially, except for the factor $(\log p)^{3/2}$, verifies Huber's conjecture. Unfortunately, Portnoy's arguments which are based on stochastic expansions are extremely delicate. The situation is somewhat easier for monotone ψ, but even there the argument is difficult.

Recently, Welsh(1987) has provided an elegant, unified approach to M-estimator asymptotics based on the stochastic equicontinuity of associated M-processes -- stochastic approximations to the defining normal equations of M-estimators. One virtue, among many, of this approach is that it yields large-p asymptotics for a somewhat larger class of M-estimators. In particular the treatment of an unknown scale parameter is treated with in this framework, as are instances of non-smooth M-estimators. In the latter category, the l_1-regression estimator and other so-called "regression quantiles" see (Koenker and Bassett(1978) and Koenker and Portnoy(1987)), are shown to be asymptotically Gaussian as $p \to \infty$ provided that $p^3(\log n)^2/n \to 0$. This is somewhat more stringent that the rates of $p^2(\log n)^{2+\gamma}/n \to 0$ for $\gamma>0$ derived by Welsh for smooth M-estimators.

While the importance of the classical linear regression model in econometrics can hardly be over-estimated, there are numerous related estimation problems which also require an asymptotic theory with parametric dimensionality tending to infinity. In a remarkable paper, Sargan(1975) addresses certain implications of large-p asymptotics in simultaneous equation models. Related results appear in Kunitomo(1981). In time-series there are numerous places where one is naturally led to sequences of models whose dimensionality tends to infinity. Hannan(1985) mentions some examples in a recent interview. Non-parametric regression in its many guises is the most obvious example: here recent work by Elbadawi, Gallant, and Souza(1983) has emphasised the centrality of the dimensionality-choice problem. Various semi-parametric models, typically involving density estimation of an infinite dimensional nuisance parameter, also require an asymptotic theory with $p \to \infty$. In short, large-p asymptotics

are an essential element of many of the current developments in econometric theory. And we are led to conclude that both the theory and practice of econometrics currently demands an asymptotic theory which explicitly considers model sequences for which $p \to \infty$ with n.

4. Epilogue

Perhaps we should pause here to reconsider some implications of the results surveyed in the previous section for the wage equation literature considered in Section 2. Recall that our empirical meta-model of wage-equations implied that p^4/n was roughly constant over the observed range of sample sizes. Thus, the foregoing results would appear to be extremely encouraging. However, we should be careful to remember that they rely on certain regularity conditions on the sequence of designs in addition to the rate conditions on the growth of p. These conditions as Portnoy shows are satisfied by design sequences drawn at random from a distribution "not too concentrated in any fixed directions." Such conditions, in a simpler form, already arise in the case of least squares where $\bar{h} \to 0$ *implied* $p/n \to 0$ as a necessary condition, but clearly the $\bar{h}$ condition, is much more stringent. For example in the p sample design it requires that the number of observations *in each cell* tends to infinity as $n \to \infty$.

Appendix

Given independent negative binomial observations, y_i, on random variables, Y_i, parameters (α_i,γ) we have log-likelihood,

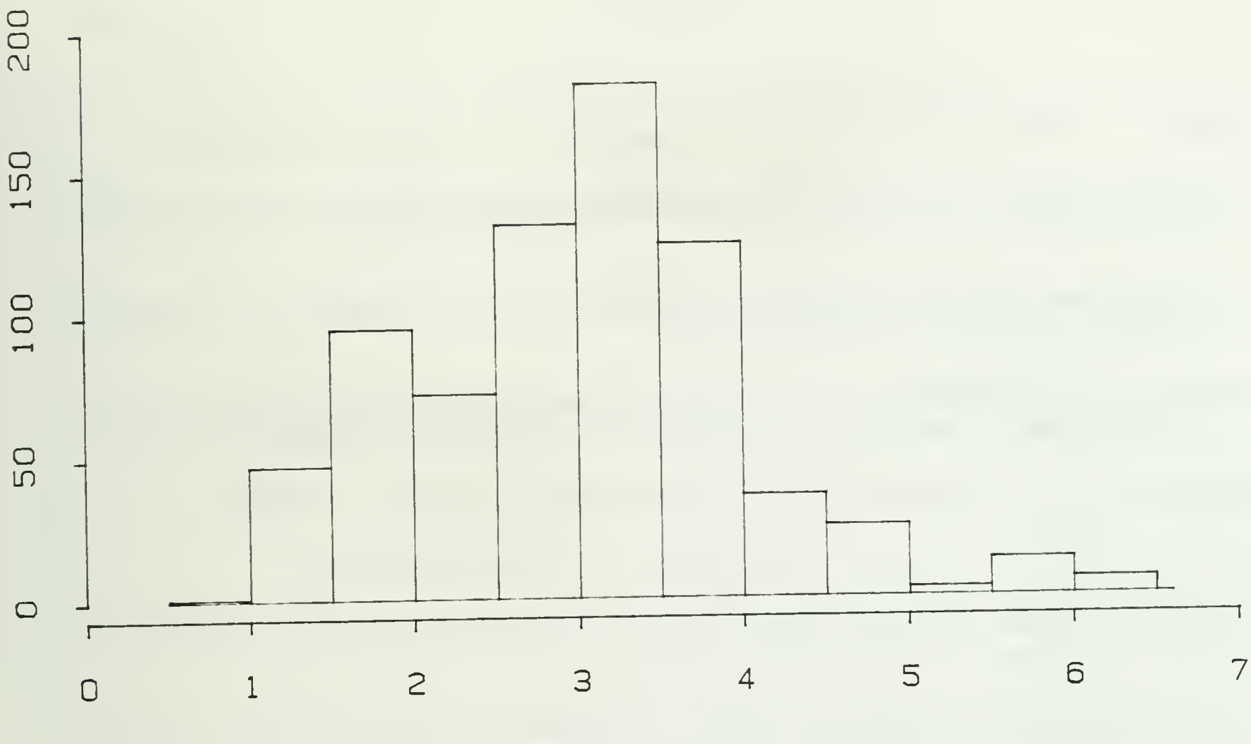
$$l(\boldsymbol{\alpha},\gamma) = \sum_{i=1}^{n}\log(\Gamma(y_i+\alpha_i))-\log\Gamma(\alpha_i)-\log\Gamma(y_i+1)+y_i\log(\gamma/(1+\gamma))-\alpha_i\log(1+\gamma)$$

In this model, $EY_i \equiv \mu_i = \alpha_i\gamma$ and $VY_i \equiv \mu_i+\mu_i^2/\gamma$ Now, if we take, $\alpha_i = \exp(x_i\beta)$ we might have for example,

$$\log EY_i = \log\gamma+\beta_1+\beta_2\log z_i$$

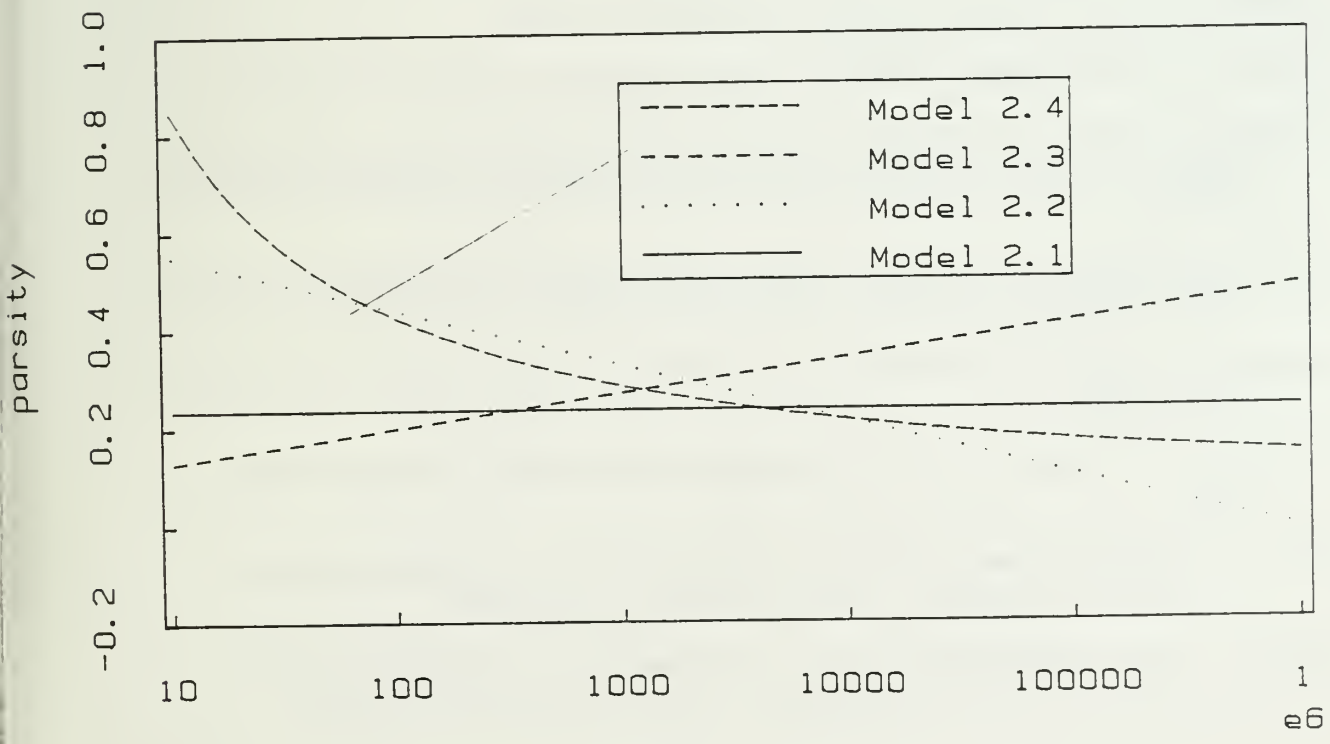
and it is straightforward to to compute elasticities from this expression. It is also clear the the variance of Y_i increases quadratically with the mean, in contrast to the Poisson model, but that as $\gamma\to\infty$ we obtain the Poisson model as a limiting case. Readers interested in a further exposition of this model and variations thereof, are urged to consult the recent survey by Trivedi and Cameron(1988). It also should be noted that misspecification of the form of the heteroscedasticity in models of this type typically leads to inconsistency of the estimator of the regression parameter. This point is explored in detail in Pagan and Sabau(1987), and may be attributed to the lack of block diagonality in the information matrix when the covariance parameters depend upon the regression parameters.

Figure 2.1



Histogram of Meta-sample Sample Sizes

Figure 2.2



Elasticity of Parsimony Functions

References

Amemiya, T. (1985). *Advanced Econometrics.* Harvard.

Anscombe, F. (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics* 5, 165-173.

Baker, R. J. and Nelder, J. A. (1978). *The GLIM System: Generalized Linear Interactive Modelling* Numerical Algorithms Group

Bassett, G. W. and Koenker, R. W. (1978). The asymptotic distribution of the least absolute error estimator. *Journal of the American Statistical Association* 73, 618-622.

Chesher, A. (1984). Testing for Neglected Heterogeneity. *Econometrica* 52, 865-872.

Cox, D. R. (1984). Some remarks on overdispersion. *Biometrika* 70, 269-274.

Elbadawi, I., Gallant, A. R., and Souza, G. (1983). An elasticity can be estimated consistently without *a priori* konwledge of functional form. *Econometrica* 51, 1731-1751.

Hausman, J., Hall B. and Griliches, Z. (1984). Econometric models for count data with an application to the patents R&D relationship. *Econometrica* 52, 909-938.

Huber, P. J. (1973). Robust regression, asymptotics, conjectures and monte-carlo. *Annals of Statistics* 1, 799-821.

Huber, P. J. (1981). *Robust Statistics.* New York: Wiley.

Johnson, N. L. and Kotz, S. (1969), *Discrete Distributions.* Wiley.

Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles. *Econometrica* 46, 33-50.

Kunitomo, N. (1981). On a third order optimum property of the LIML estimator when the sample size is large. Technical report of the Department of Economics, Northwestern University.

Lai, T. L., Robbins, H. and Wei, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proceedings of the National Academy*, 75, 3034-3036.

Lancaster, A. B. (1984). The covariance matrix of the information matrix test. *Econometrica* 52, 1051-1054.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* London: Chapman and Hall.

Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I: Consistency. *Annals of Statistics* 12, 1298-1309.

Portnoy, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II: Normal Approximation. *Annals of Statistics* 13, 1403-1417.

Sargan, J. D. (1975). Asymptotic Theory and Large Models, *International Economic Review*, 16, 75-91.

Spady, R. H., (1984). QMLE: A program for quasi-maximum likelihood estimation, Bell Communications Research.

Welsh, A. H. (1987). "On M-Processes and M-Estimation," Technical Report 213, Department of Statistics, University of Chicago.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1-25.

Yohai, V. and Marrona, R. A. (1979). Asymptotic behavior of M-estimators for linear models. *Annals of Statistics* 7, 258-268.